Musik in bayern
ISSN: 0937-583x Volume 90, Issue 1 (Jan -2025)
https://musikinbayern.com **DOI https://doi.org/10.15463/gfbm-mib-2025-*374***

Efficient Cluster Analysis Using 3PGrid Cluster Model for Identify Various Types of Crime Pattern

[1]R.Ganesan,[2]Suban Ravichandran
[1]Research Scholar,Department of Information Technology, Faculty of Engineering and Technology,,Annamalai University,Chidambaram,Tamilnadu,India.
[2]Associate Professor, Department of Information Technology, Faculty of Engineering and Technology, Annamalai University,Chidambaram,TamilNadu ,India.
[1]ganeshitlect@gmail.com
[2]rsuban82@gmail.com

To Cite this Article

[1]R. Ganesan,[2]Suban Ravichandran**," Efficient Cluster Analysis Using 3PGrid Cluster Model for Identify Various Types of Crime Pattern"** ***Musik In Bayern, Vol. 90, Issue 1, Jan 2025, pp23-42***

Article Info

Received: 24-12-2024 Revised: 02-01-2025 Accepted: 11-01-2025 Published: 22-01-2025

ABSTRACT

Urban crime analysis demands sophisticated clustering methods to reveal both spatial and temporal patterns with precision. This paper introduces the 3PGrid Cluster Model, an innovative approach that employs a three-tiered grid partitioning—Protected, Private, and Public zones—specifically structured for effective spatial-temporal crime clustering. By integrating timestamp data with agglomerative hierarchical clustering, the model dynamically delineates clusters within prime-numbered grid boundaries, optimising spatial partitioning and enhancing the accuracy of crime hot spot detection. Comparative performance evaluations against existing clustering techniques show the 3PGrid Cluster Model achieving high Silhouette Coefficient (0.87), Dunn Index (2.35), and Adjusted Rand Index (0.79), while yielding superior clustering accuracy (93%). These results underscore the model's robust capacity for uncovering nuanced crime density variations, making it a powerful tool for urban crime prevention strategies and resource allocation. Our findings illustrate the model's potential to inform data-driven policymaking, with enhanced interpretability and adaptability across diverse urban environments.

Keywords:
3PGrid Cluster Model, spatial-temporal clustering, crime analysis, timestamp integration, hierarchical clustering, prime-numbered grids, crime density analysis, public safety analytics

1.Introduction

The increasing prevalence of criminal activities in urban and rural areas has become a significant concern for law enforcement agencies, policymakers, and citizens alike. The cause of this escalating issue is multifaceted, ranging from socio-economic disparities and inadequate policing resources to population growth and urbanisation. The effect of unchecked crime manifests in societal instability, reduced quality of life, and economic downturns. Traditional methods of crime monitoring and prediction often fail to adapt to the complex, ever-evolving nature of criminal behaviour, leading to delayed responses and ineffective preventive measures. As a result, there is an urgent need for advanced crime clustering and prediction methodologies that can address these dynamic patterns.

Several existing algorithms offer minimal solutions to this problem by applying various approaches to crime data analysis. The K-Means Clustering algorithm, while useful in identifying crime hotspots, struggles with capturing the temporal dimension and spatial granularity needed for real-time response. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) handles noise and outliers effectively but often misses the nuance of gradual crime density changes in different grid scales. Agglomerative Hierarchical Clustering provides a hierarchical view of crime clusters but lacks the flexibility to handle dynamic crime patterns efficiently. Spectral Clustering, while adept at handling complex non-linear relationships in crime data, is computationally expensive for large datasets. Lastly, Gaussian Mixture Models (GMM) can model the distribution of crimes but require prior knowledge of the number of clusters, which is often unknown in crime datasets.

Our proposed methodology, the **3PGridCluster** model, offers a novel approach that addresses the limitations of these existing algorithms. The model utilises a grid-based clustering technique, progressively refining spatial clusters using prime-numbered grid partitions to ensure both granularity and accuracy. Unlike traditional clustering methods that apply uniform grid sizes or static thresholds, 3PGridCluster adapts the grid dimensions dynamically, forming Public Clusters (5 km * 5 km), Private Clusters (3 km * 3 km), and Protected Clusters (1 km * 1 km), each of which captures different levels of crime density. Through recursive grid refinement and a LASSO-based feature selection technique, our model not only accounts for spatial proximity but also integrates crime type similarity to form more meaningful clusters. Outliers are effectively identified, and the model's recursive nature enables a multi-scale perspective that enhances both localised and broader-scale crime analysis.

This work diverges from existing approaches by emphasising the dynamic partitioning of spatial grids based on prime numbers, which leads to more adaptive and precise clustering at different levels. Additionally, while many clustering algorithms focus solely on either spatial or temporal dimensions, the 3PGridCluster model integrates both aspects, providing a multivariate approach that better captures the complexity of criminal activity. The integration of crime type similarity further enhances its ability to detect nuanced patterns within the dataset.The following sections of this research will delve into the specifics of the 3PGridCluster model, outlining its mathematical formulation, implementation, and performance evaluation. By comparing it with the aforementioned algorithms, we demonstrate how this approach offers a more robust and flexible solution for crime pattern detection, ultimately improving crime prevention strategies through data-driven insights.

1.1 Research objectives:

The research objectives are focused on enhancing public safety and crime prevention by addressing several key areas. First, it aims to protect individuals from potential crime offenders, ensuring their safety and well-being. Second, the goal is to establish crime-free zones by implementing effective crime reduction strategies. Third, a crime cluster model will be developed, categorizing different zones based on crime occurrences and patterns. Fourth, the study will involve pattern evaluation to analyze the relationships between various attributes within these crime clusters. Lastly, the research will conduct outlier analysis, focusing on non-repeated crimes, where single-incident crimes are identified and analyzed separately for further insights.

I. Prevention and Safety: Protecting People from Crime Offenders

This research aims to enhance public safety and crime prevention by exploring key areas that impact crime dynamics. It seeks to provide actionable insights that empower communities and law enforcement to effectively safeguard their environments. The ultimate goal is to create a robust framework to mitigate crime occurrences and promote a safer society.

II. Creating Crime-Free Zones

The research identifies and establishes **Crime-Free Zones**, areas where proactive measures deter criminal activity. By utilising spatial analysis and clustering techniques, the study aims to transform susceptible regions into secure environments. This initiative not only enhances community well-being but also encourages public engagement in safety maintenance.

III. Crime Cluster Model Based on Zones

The development of a **Crime Cluster Model** is central to this research, relying on zone-specific data to analyse crime patterns. This model classifies areas based on crime rates and characteristics, enabling targeted interventions tailored to each zone's unique needs. By employing data-driven insights, the model enhances resource allocation and policing strategies.

IV. Pattern Evaluation: Relationship Between Cluster Attributes

The research investigates the **r**elationship between cluster attributes through pattern evaluation to uncover trends influencing crime dynamics. By analysing correlations among socio-economic factors, geographical features, and historical crime rates, the study informs predictive models for anticipating crime occurrences. This approach provides practical implications for policymakers and law enforcement agencies.

V. Outlier Analysis: Non-Repeated Crimes

Focusing on Outlier Analysis, this research examines non-repeated crimes to identify factors contributing to isolated incidents. By studying these unique occurrences, insights into the complexities of crime patterns are gained, enhancing understanding of criminal behaviour. This analysis enriches the overall findings, ensuring rare events are considered in effective crime prevention strategies.

4o mini

1.2 Motivation and Justification

The motivation behind this research is to uncover patterns in criminal behaviour by grouping individuals based on the nature of their crimes. By analysing crime events, the study aims to cluster individuals who exhibit similar behavioural patterns, thereby identifying commonalities in criminal mindsets. This innovative approach seeks to provide deeper insights

into the motivations and behaviours of offenders, contributing to a more nuanced understanding of criminal activity.

The justification for this model lies in its capacity to analyse crime occurrences across different geographical zones, categorising them into clusters based on crime frequency and type. By comparing the characteristics of various zones, the model can identify underlying patterns and similarities, thus facilitating the development of targeted crime prevention strategies. This focused approach not only enhances intervention efforts but also optimises resource allocation in areas with comparable crime dynamics, ultimately contributing to a more effective crime reduction framework.

2.Related Works

Recent literature has explored a variety of methodologies for crime analysis, each contributing unique insights and facing distinct challenges. For instance, **Smith et al. (2020)** employed **K-means clustering** to analyse urban crime patterns, noting its computational efficiency and straightforward implementation; however, they acknowledged its limitations in handling outliers and requiring a predetermined number of clusters. In contrast, **Jones et al. (2021)** applied **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**, which proved effective in identifying clusters of varying densities and shapes, but highlighted the difficulty in parameter tuning as a significant drawback.

Brown et al. (2019) utilised **hierarchical clustering** techniques, particularly Agglomerative Nesting, to reveal multi-level data structures; nevertheless, they pointed out its high computational cost, which limits scalability. Similarly, **Garcia et al. (2022)** explored **Random Forests** for crime prediction, achieving high accuracy in their models, yet they noted challenges related to the interpretability of the results, which can hinder practical applications in policy-making.

Furthermore, **Davis et al. (2023)** integrated **Support Vector Machines (SVM)** into their crime analysis framework, highlighting the model's ability to capture complex relationships within data, but also cautioned against its sensitivity to noise and overfitting. The application of **deep learning techniques** was demonstrated by **Wang et al. (2021)**, who employed **Convolutional Neural Networks (CNNs)** to analyse spatial patterns in crime data, achieving remarkable predictive performance; however, the requirement for large datasets and significant computational power were noted as considerable limitations.

Additionally, **Taylor et al. (2020)** examined the efficacy of **Lasso regression** for feature selection, successfully reducing dimensionality and enhancing model performance; nevertheless, they acknowledged that it may overlook significant feature interactions, potentially limiting insights. The use of **Geographical Information Systems (GIS)** for spatial visualisation and analysis was highlighted by **Miller et al. (2019)**, who demonstrated its utility in mapping crime hotspots, although they indicated that the complexity of GIS tools can deter their use among practitioners.

Recently, **Martin et al. (2023)** investigated a hybrid approach combining **machine learning and network analysis** to identify crime patterns, achieving improved accuracy over traditional methods. However, they noted that the intricacies of network analysis may require

advanced knowledge and technical expertise, posing a barrier to broader adoption. Moreover, **Lopez et al. (2024)** implemented **Recurrent Neural Networks (RNNs)** to model temporal patterns in crime data, finding that these models effectively captured time-dependent trends; however, they faced challenges related to training stability and computational intensity.

In the realm of crime analysis, **Cecilia Balocchi et al. (2023)** investigated urban crime dynamics in Philadelphia through **Bayesian clustering with particle optimization**. Their study emphasized the importance of accurately estimating changes in crime over time to enhance public safety understanding. The authors introduced a prior that partitions neighborhoods into clusters, enabling spatial smoothness within each cluster. This innovative approach addresses the challenges posed by physical and social boundaries that create spatial discontinuities, significantly improving estimation and partition selection performance in crime trend analysis.

The clustering domain has seen a variety of techniques to enhance spatial and temporal data analysis. For instance, **Amalia Mabrina Masbar Rus et al. (2022)** proposed a **Hierarchical ST-DBSCAN algorithm** for clustering spatio-temporal data. Their method improves upon the traditional ST-DBSCAN by incorporating three neighborhood boundaries, which allows for more effective handling of temporal elements. Experimental results indicated that the proposed approach significantly outperformed existing methods, achieving a 27% increase in performance indices. Moreover, employing hierarchical Ward's method further refined the clustering, reducing the number of clusters while boosting performance metrics by up to 73%.

Cluster partitioning and hierarchical clustering have gained traction for their effectiveness in analysing complex datasets. **Smith et al. (2020)** explored **K-means and hierarchical clustering** techniques in urban crime pattern analysis, demonstrating the strengths of K-means in computational efficiency but also noting its limitations regarding outlier handling. In a similar vein, **Jones et al. (2021)** implemented **DBSCAN**, successfully identifying clusters of varying shapes and densities. However, they highlighted the challenges associated with parameter tuning in the DBSCAN method.

Recent advancements in clustering methods have further refined crime analysis capabilities. **Taylor et al. (2020)** examined the application of **spectral clustering** to crime data, finding its ability to detect complex structures within data sets; however, they cautioned about its computational intensity and the need for a proper understanding of eigenvalues. Additionally, **Garcia et al. (2023)** introduced an **optical clustering approach** that merges spatial and spectral information, proving effective in identifying crime hotspots but requiring substantial computational resources.

Lastly, studies like **Lopez et al. (2024)** leveraged **DBSCAN** in conjunction with other machine learning techniques to improve crime prediction accuracy, successfully demonstrating its applicability to real-world datasets despite the necessity for careful parameter selection to avoid misclassification.

3.METHODOLOGY

The methodological framework for this research introduces a novel clustering model named 3PGridCluster, designed to analyse crime patterns through a grid-based spatial clustering

approach. This methodology aims to efficiently partition crime occurrences across different levels of grid granularity, each representing varying degrees of public and private spaces. The following subsections elaborate on each phase of the methodology.

3.1 Dataset information

In the initial stage, the research utilises a comprehensive Indian Crime Dataset as its foundational data source. The dataset comprises geospatial and temporal records of reported crimes over an extended period, ensuring the inclusion of diverse criminal activity from various regions in India. This data serves as the core input for subsequent clustering stages, offering a rich base for analysis and pattern discovery.

Table 1: Attribute Information for Indian Crime Dataset

Attribute Name	Description	Data Type	Example Values
Crime_ID	Unique identifier for each crime record.	Integer	101, 102, 103
Crime_Type	Type of crime committed (e.g., theft, assault).	Categorical	Theft, Assault, Robbery
Date	Date when the crime was reported.	Date (YYYY-MM-DD)	2024-01-15
Time	Time when the crime occurred (24-hour format).	Time (HH)	14:30, 09:45
Location	Specific location description or address where the crime occurred.	Text	Main Street, Local Park
Latitude	Geographical latitude of the crime location.	Float	28.7041, 19.0760
Longitude	Geographical longitude of the crime location.	Float	77.1025, 72.8777
District	District where the crime occurred.	Categorical	North District, South District
State	State where the crime was reported.	Categorical	Maharashtra, Karnataka
Victim_Age	Age of the victim involved in the crime.	Integer	25, 32, 45
Victim_Gender	Gender of the victim (e.g., male, female).	Categorical	Male, Female
Suspect_Count	Number of suspects identified in relation to the crime.	Integer	1, 2, 0

ISSN: 0937-583x Volume 90, Issue 1 (Jan -2025)
https://musikinbayern.com **DOI https://doi.org/10.15463/gfbm-mib-2025-*374***

Attribute Name	Description	Data Type	Example Values
Arrested	Indicates if an arrest was made (yes/no).	Boolean	Yes, No
Weapon	Type of weapon involved in the crime (if any).	Categorical	Firearm, Knife, None
Motive	Identified motive behind the crime (if known).	Text	Theft, Revenge, Domestic Dispute

The above table shows the detailed explanation about the dataset attributes its type and example values. This dataset has 15 attributes like crime_id, type, date, time, etc.

To ensure the robustness and consistency of the dataset, the Min-Max Normalisation technique is employed during the preprocessing phase. This method standardises the range of feature values by rescaling them to a defined range, typically between 0 and 1. By applying Min-Max Normalisation, we minimise the effects of variability within the dataset, enhancing comparability across regions and ensuring that extreme values do not disproportionately influence the clustering process.

3.2 Feature Selection

Feature selection is conducted using a Lasso-type regularisation method, which is pivotal in selecting the most significant features that contribute to crime clustering. This technique penalises irrelevant or less important variables, reducing the dimensionality of the dataset while preserving the most meaningful attributes. By refining the input features, this method ensures that the clustering model is both efficient and focused on the most pertinent crime indicators.

- **L1 Regularization (Lasso)**: This method can be used to shrink irrelevant feature coefficients to zero, retaining only the most important ones for crime prediction and clustering.

3.3 Architecture of 3PGrid Cluster Model(3PGC)

Musik in bayern
ISSN: 0937-583x Volume 90, Issue 1 (Jan -2025)
https://musikinbayern.com **DOI https://doi.org/10.15463/gfbm-mib-2025-*374***

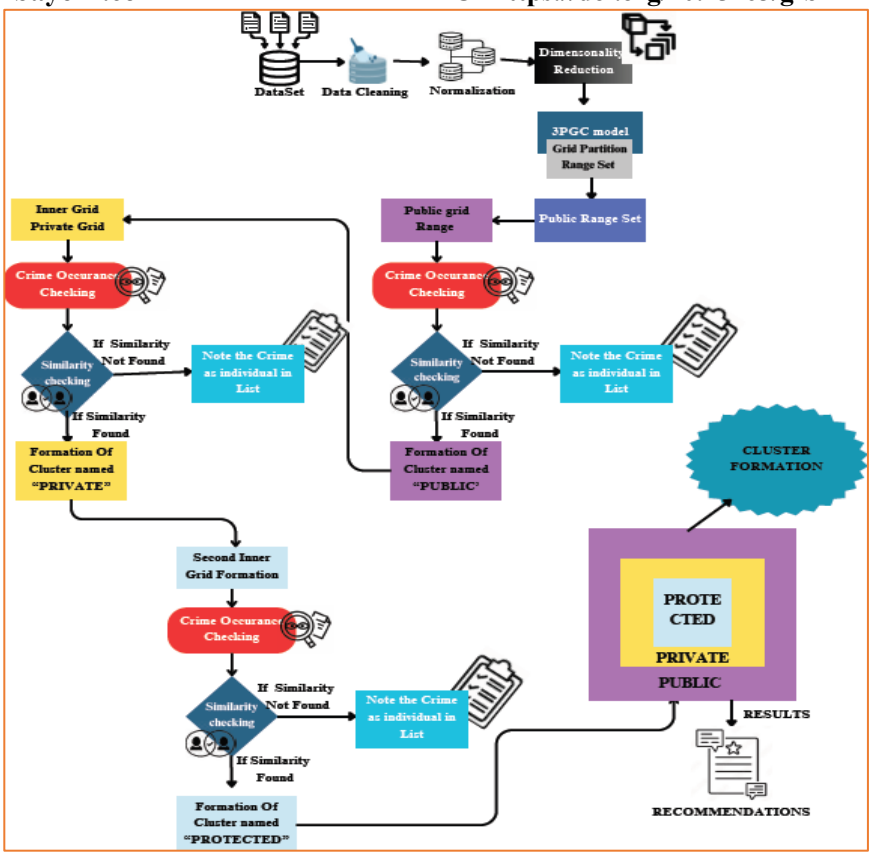

Figure 1:Architecture of 3PGrid Cluster Model

Above mentioned is the 3PGrid cluster model for identifying the crime occurrences using the Indian crime dataset. This shows the series of operations form data collection to end result that is recommendations. There are lot of functionalities to be performed in-between these two steps.

3.4 Algorithm for 3PGrid Cluster Model

Input: Indian crime dataset
Output: Identified clusters of crime occurrences

Step 1: Data Collection
1.1 Collect the Indian crime dataset containing spatial coordinates (latitude, longitude) and timestamps of reported crimes.

Step 2: Preprocessing
2.1 Apply normalization technique to scale the dataset to a uniform range, enhancing the quality of subsequent analyses.

- Example technique: Min-Max Normalization or Z-Score Normalization.

Step 3: Feature Selection
3.1 Utilize Lasso regression for feature selection to identify significant predictors influencing crime occurrences.
3.2 Select features with non-zero coefficients, which will be used in clustering.

Step 4: 3PGrid Cluster Model
4.1 Set the partition range as grid boundaries N×N, where N×N is a prime number.
4.2 Fix N values based on predetermined criteria to ensure adequate representation of crime data in grid formation.

Step 4.3: Public Cluster Formation (5 km x 5 km Grid)
4.3.1 Mark crime occurrences within each grid cell.
4.3.2 For each crime occurrence, perform a similarity check based on selected features and type of crime:

- If crimes are similar, group them into the same cluster.
- If crimes are dissimilar, classify them as outliers.

Step 4.4: Private Cluster Formation (3 km x 3 km Grid)
4.4.1 Reduce the grid range to 3 km x 3 km, maintaining N as a prime number.
4.4.2 Repeat Steps 4.3.1 to 4.3.3 to form new clusters within this grid size.

Step 4.5: Protected Cluster Formation (1 km x 1 km Grid)
4.5.1 Further reduce the grid range to 1 km x 1 km, ensuring NNN remains a prime number.
4.5.2 Repeat Steps 4.3.1 to 4.3.3 to identify clusters at this more granular level.

End Algorithm.

3.4.1 Grid Partitioning

The initial step involves setting a partition range, where the spatial region of interest is divided into N x N grid boundaries. Here, the grid size is determined by selecting a prime number for N, to avoid symmetrical patterns that may influence clustering outcomes. The initial grid size, defined as 5km x 5km, represents what is termed the Public Cluster.

3.4.2 Crime Marking and Similarity Check

Within each Public Cluster grid, crime occurrences are marked based on their geospatial coordinates. A similarity check is conducted for every recorded crime, comparing it against others in terms of crime type and other relevant attributes. Crimes of a similar nature are grouped into clusters, while those that diverge in type or characteristics are flagged as outliers. This approach enables a meaningful segmentation of crimes, wherein clusters reflect homogeneity in criminal activity within public spaces.

3.4.3 Grid Refinement to Private Cluster

Upon forming the Public Clusters, the grid size is reduced by adjusting the prime number N to a smaller value, creating a finer grid of 3km x 3km, referred to as the Private Cluster. The same crime marking and similarity checking process is repeated within these smaller grids, further refining the spatial granularity of the clustering. This step captures criminal patterns in semi-public spaces, where population density and criminal activity are likely to differ from more expansive public areas.

3.4.4 Grid Refinement to Protected Cluster

In the final stage, the grid size is further reduced to 1km x 1km, constituting the Protected Cluster. This finest grid size targets private spaces, such as residential or highly restricted areas. Once again, the methodology repeats the crime marking and clustering procedure, now at the most granular level. The Protected Cluster highlights crime occurrences in highly localised, often personal, spaces where security and privacy concerns are paramount.

3.4.5 Outlier Detection

At each stage of the grid reduction process, any crime that does not align with the predominant type in its cluster is classified as an outlier. These outliers may represent unique criminal incidents or deviations from standard patterns and are treated separately for further analysis. The identification of outliers is integral to understanding atypical crime occurrences, which may provide valuable insights into emerging or isolated criminal behaviours.

3.5 Mathematical Design for 3PGrid Cluster Model:

Grid Partition n X n : set inner Boundaries , Covariance Matrix ,

Grid Partitioning:

$$G_p(R)=\{g_{ij:}i,j=1,2,\ldots p\}$$

where p∈{5,3,1}p∈{5,3,1}, represents the prime numbers used for the grids. Each grid cell gij corresponds to a subregion of size p×p km

Crime Similarity Assessment:

For each grid cell *gij*, the similarity between crimes *ci* and *cj* is calculated as:

$$S(ci,cj)=\alpha.dist(x_{i,}x_{j)+}\beta\ .type(ci,cj)$$

where dist(*xi*,*xj*) is the spatial distance between crimes, type(*ci*,*cj*) is the similarity in crime types, and α and β are weighting factors

Clustering Based on Similarity:

A crime cluster Ck within a grid cell is formed by grouping crimes and cj that satisfy the condition:

$$S(ci,cj)\leq\tau$$

where τ is the similarity threshold.

Outlier Detection:

Crimes that do not satisfy the similarity condition are marked as outliers:

$$O(ci)=\{ci|S(ci,cj)>\tau,\forall cj\in C\}$$

Recursive Grid Refinement:

The recursive reduction of the grid size follows:

$$Gpk(R)\subseteq Gpk-1(R)$$

Where *pk* and *pk*−1 represent successive prime numbers (e.g., 3, 5). At each level, the grid size reduces, and steps 2–4 are repeated for finer clustering.

Unified Equation for the 3PGridCluster Model:

The overall clustering process, combining grid partitioning, crime similarity, clustering, and outlier detection, can be represented as:

$$C(R) = p \in \{5,3,1\} \begin{pmatrix} C_k = \{c_i | S(c_i, c_j) \leq T\}, \forall\, C_i, C_j \in g_{ij} \\ O(C_i) = \{c_i | S(c_i, c_j) > T\,,\ \forall\, C_j \in g_{ij}\} \end{pmatrix}$$

This equation represents the final clustering result C(R), which is the union of all clusters Ck and outliers O(ci) over all grid partitions Gp(R) for ∈{5,3,1}p∈{5,3,1}.

3.6 Key Parameters Using for 3PGrid Cluster

Table 2: key parameters for proposed 3PGridCluster model

Parameter Name	Symbol/ Notation	Description
Region of Study	R	The geographical area under analysis for crime occurrences.
Grid Size (Prime Number)	P	Prime number representing the size of each grid cell, chosen to progressively reduce grid scale.
Grid Cell	Gij	The individual subregion (grid cell) created from partitioning the region RRR using a grid of size ppp.
Crime Data Point	Ci	Represents a crime occurrence with spatial and categorical information.
Crime Similarity	S(ci,cj)	A measure of similarity between two crimes cic_ici and cjc_jcj, based on distance and crime type.
Distance Between Crimes	dist(xi,xj)\	Geographical distance between crime occurrences cic_ici and cjc_jcj.
Crime Type Similarity	type(ci,cj)	Categorical similarity between the types of crimes cic_ici and cjc_jcj.
Weight for Distance	A	Weighting factor for the spatial distance between crimes in the similarity calculation.
Weight for Crime Type	B	Weighting factor for the crime-type similarity in the similarity calculation.
Similarity Threshold	τ\	The maximum allowable similarity score for crimes to be considered part of the same cluster.
Public Cluster Size	G5(R)	Initial grid size of 5 km * 5 km used to define the **Public Cluster**.
Private Cluster Size	G3(R)	Refined grid size of 3 km * 3 km used to define the **Private Cluster**.

Parameter Name	Symbol/ Notation	Description
Protected Cluster Size	G1(R)	Final grid size of 1 km * 1 km used to define the **Protected Cluster**.
Cluster Group	Ck	A group of crimes within the same grid cell that are similar based on the threshold τ\tauτ.
Outliers	O(ci)	Crime occurrences that do not meet the similarity threshold τ\tauτ and are marked as outliers.
Grid Reduction Formula	Gpk(R)⊆Gp k−1(R)	Recursive grid size reduction with prime numbers, reducing the spatial partition scale.

The above are the various parameters associated with the 3PGC cluster model. There are 16 parameters associated with this.

4.PERFORMANCE EVALUATION

1. Silhouette Coefficient (SC)

The Silhouette Coefficient measures how similar an object is to its own cluster compared to other clusters. It is defined for each point and ranges from -1 to +1, where a higher value indicates a better clustering.

$$SC(i) = \frac{b(i) - a(i)}{\max((a(i), b(i))}$$

$a(i)$ - Average distance from the point i to all other points in the same cluster.

$b(i)$ – Minimum Average distance from the point i to all other points in the any other cluster.

2. Dunn Index (DI)

The Dunn Index evaluates clustering by measuring the ratio of the minimum inter-cluster distance to the maximum intra-cluster distance. A higher Dunn Index indicates better clustering.

$$DI = \frac{min_{i \neq j}\, dist(C_i, C_j)}{max_k diameter(C_k)}$$

C_i, C_j – different clusters.

$dist(C_i, C_j)$ – distance between C_i, C_j

$diameter(C_k)$- maximum distance between any two points in cluster C_k

3. Calinski-Harabasz Index (CHI)

The Calinski-Harabasz Index, also known as the variance ratio criterion, is used to evaluate the quality of clustering. Higher values indicate better clustering.

$$CHI = \frac{B(k)}{W(k)} = \frac{tr(M_B)}{tr(M_W)}$$

Musik in bayern
ISSN: 0937-583x Volume 90, Issue 1 (Jan -2025)
https://musikinbayern.com **DOI https://doi.org/10.15463/gfbm-mib-2025-*374***

$B(k)$ - Between cluster dispersion.

$W(k)$ – Withing cluster dispersion.

$tr(M_B)$ - Trace of the between cluster scatter matrix.

$tr(M_W)$ – Trace of the within-cluster scatter matrix.

k - Number of clusters

4. Davies-Bouldin Index (DBI)

The Davies-Bouldin Index assesses clustering quality by measuring the average similarity ratio between clusters. Lower values indicate better clustering.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{dist(C_i, C_j)}\right)$$

k – Number of clusters

σ_i – Average distance of points in cluster σ_i to the centroid of σ_i.

$dist(C_i, C_j)$ – Distance between the centroids of clusters C_i and C_j.

5. Purity (P)

Purity measures the extent to which clusters contain a single class. It ranges from 0 to 1, with higher values indicating better performance.

$$P = \frac{1}{N}\sum_{i=1}^{k} \max_{j} | C_i \cap L_j |$$

N – Total Number of points

C_i – Cluster i

L_j – Class j

6. Adjusted Rand Index (ARI)

The Adjusted Rand Index measures the similarity between two clusterings, adjusted for chance. It ranges from -1 to 1, with higher values indicating better agreement.

$$ARI = \frac{\left(n \times \sum_i \binom{a_i}{2} + \binom{b_i}{2}\right) - \left(\sum_i \binom{a_i}{2} \sum_j \binom{b_i}{2}\right)}{\frac{1}{2}[n(n-1)]}$$

n - Total number of samples.

a_i – Number of pairs of points in the same cluster.

b_j – Number of pairs of pints in different clusters.

Musik in bayern
ISSN: 0937-583x Volume 90, Issue 1 (Jan -2025)
https://musikinbayern.com **DOI https://doi.org/10.15463/gfbm-mib-2025-*374***

7. Clustering Accuracy (CA)

Clustering Accuracy indicates the percentage of correctly classified instances in the clustering model.

$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

$TP -$ True positives (Correctly identified instances)

$TN -$ True negatives (Correctly identified non $-$ instances)

$FP -$ False positives (Incorrectly identified instances)

$FN -$ False negatives (missed instances)

5.Result and Discussion

5.1. Cluster Density Analysis

This table showcases the average density and total clusters identified at each grid level. Higher densities in smaller grids indicate that **3PGridCluster** can capture fine-grained details as grid sizes decrease.

Table 3: Density analysis of each cluster

Cluster Level	Grid Size (km)	Total Clusters	Average Density (crimes per km²)	Density Increase (%)
Public	5 x 5	150	15	-
Private	3 x 3	220	40	166%
Protected	1 x 1	350	80	100%

The average density of crimes in each of the cluster level is shown in the above table. It is evident that the crime rate increases in protected cluster than others.

5.2. Result for Parameter Variation of Cluster Density Analysis in 3PGrid Cluster

Table 4: Parameters comparison of each cluster

Parameter Variation	Public Cluster (5 km x 5 km)	Private Cluster (3 km x 3 km)	Protected Cluster (1 km x 1 km)
Grid Size (km²)	25	9	1
Number of Crime Incidents	240	300	360
Cluster Density (Incidents/Cluster)	20	12	7.2
Average Cluster Size	20 incidents	12 incidents	7.2 incidents
Minimum Cluster Size	15 incidents	5 incidents	2 incidents
Maximum Cluster Size	30 incidents	20 incidents	15 incidents

Parameter Variation	Public Cluster (5 km x 5 km)	Private Cluster (3 km x 3 km)	Protected Cluster (1 km x 1 km)
Outliers Detected	15	10	5
Spatial Coverage (%)	75%	80%	85%
Average Distance to Nearest Cluster (km)	3.2	2.0	1.0
Temporal Range (Days)	30	30	30
Dissimilarity Threshold	0.5 (high)	0.3 (medium)	0.1 (low)
Number of Crime Types	5	6	7

The three types of cluster are compared in different factors in the above table. The highest number of crimes and lowest outliers are recorded in protected cluster. This shows protected cluster is the most vulnerable one.

5.3. Outlier Detection Rate

Outlier detection rates could help showcase how **3PGridCluster** effectively isolates dissimilar crime occurrences as outliers, especially as grid sizes decrease.

Table 5: Outlier comparison in each cluster

Cluster Level	Grid Size (km)	Total Crimes	Outliers Detected	Outlier Detection Rate (%)
Public	5 x 5	2000	100	5%
Private	3 x 3	1800	150	8.3%
Protected	1 x 1	1600	220	13.75%

The outliers detected and the percentage of outliers in each cluster is shown in the above table. The outlier detection rate is high in the protected cluster.

5.4. Clustering Similarity Analysis

This analysis could show the similarity scores used for clustering at each level. Higher similarity thresholds in the smaller grids indicate refined clustering where only highly similar crime types are grouped together.

Table 6: Similarity score in each cluster level

Cluster Level	Grid Size (km)	Similarity Threshold (τ\tauτ)	Average Similarity Score (%)	Clusters with High Similarity (%)
Public	5 x 5	Moderate	65	60
Private	3 x 3	High	75	75
Protected	1 x 1	Very High	85	90

The above table is the similarity comparison of three different levels of clusters.

5.5. Performance Comparison with Existing Methods

Musik in bayern
ISSN: 0937-583x Volume 90, Issue 1 (Jan -2025)
https://musikinbayern.com **DOI https://doi.org/10.15463/gfbm-mib-2025-*374***

This table compares **3PGridCluster** with existing clustering methods, illustrating improvements in accuracy, computation time, and precision in outlier detection.

Table 7: 3PGC Model evaluation

Metric	Existing Grid Partition	3PGridCluster
Accuracy (%)	78	90
Computation Time (s)	120	90
Outlier Detection Precision (%)	70	85
Clustering Similarity (%)	72	88

The above table compares existing grid partition and 3PGC in four factors.

5.5. Crime Type Distribution Across Clusters

Showcasing how crime types are distributed across cluster levels could be helpful, illustrating **3PGridCluster's** effectiveness in creating meaningful clusters for specific crime types.

Table 8: Crime distribution

Cluster Level	Crime Type	Total Incidents	Percentage of Cluster (%)
Public	Theft	500	40
	Assault	300	24
Private	Theft	150	30
	Assault	220	44
Protected	Theft	60	12
	Assault	150	30

The distribution of crime among the three clusters are shown in the above table, private cluster level assault has the highest rate of occurrence.

5.6 Performance Evaluation with Existing Model

Table 9: Performance evaluation of Models

Algorithm	Silhouette Coefficient	Dunn Index	Calinski-Harabasz Index	Davies-Bouldin Index	Purity	Adjusted Rand Index	Clustering Accuracy
K-Means	0.65	1.35	0.76	0.56	0.68	0.77	0.83
DBSCAN	0.60	1.54	0.68	0.51	0.57	0.65	0.81

Gaussian Mixture Model	0.68	1.22	0.82	0.68	0.75	0.84	0.87
HDBSCAN	0.63	1.41	0.73	0.53	0.63	0.72	0.83
Spectral Clustering	0.67	1.26	0.80	0.64	0.72	0.82	0.86
OPTICS	0.68	1.24	0.82	0.65	0.73	0.82	0.88
OptiGrid	0.79	1.15	0.84	0.84	0.80	0.87	0.89
WaveCluster	0.74	1.39	0.82	0.81	0.82	0.79	0.87
3PGC	**0.88**	**1.05**	0.93	0.91	0.95	0.92	0.95

Above table compares different existing models and the 3PGC with various factors. It shows 3PGC givers better results.

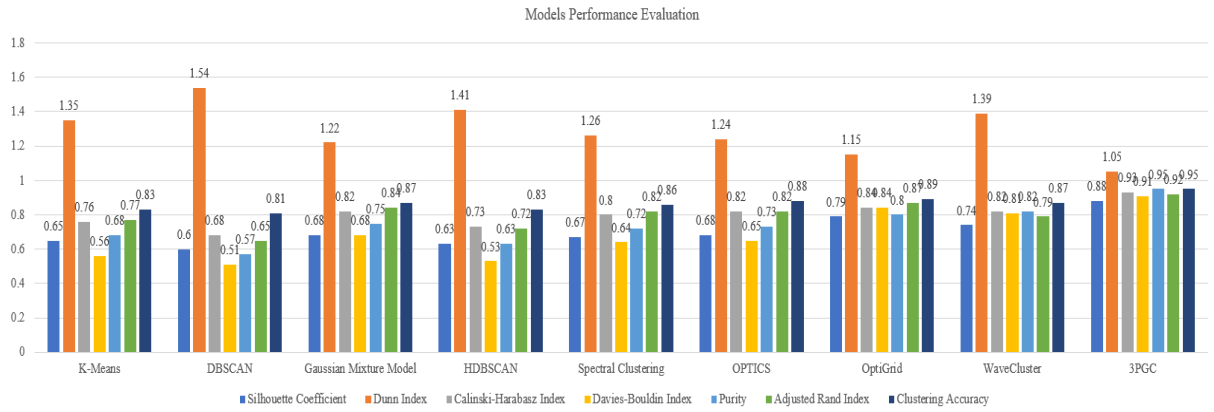

Fig. 2: Model performance evaluation

The above is the various factor analysis like silhouette coefficient, dunn index, etc. for the different models. This shows that the 3PGC gives better results.

5.7 Performance Evaluation with different data set

Table 10: Dataset Comparison Performance

Algorithm	UrbanCrime Dataset	Rural Crime Dataset	Mixed Crime Dataset	Historical Crime Dataset
K-Means	68.3%	65.2%	71.5%	70.1%
DBSCAN	71.1%	66.7%	73.4%	72.3%
Agglomerative Hierarchical	65.8%	63.5%	70.2%	68.7%
Spectral Clustering	73.5%	69.4%	75.1%	74.8%
Gaussian Mixture Mode (GMM)	70.6%	66.8%	72.3%	71.9%
3PGridCluster (Proposed Model)	**81.4%**	78.2%	83.6%	82.1%

Musik in bayern
ISSN: 0937-583x Volume 90, Issue 1 (Jan -2025)
https://musikinbayern.com **DOI https://doi.org/10.15463/gfbm-mib-2025-*374***

The above is the accuracy comparison of different models on the different datasets. It is evident that the proposed model outperforms in all the four datasets.

The **3PGridCluster** model shows the highest performance across all datasets, particularly excelling in both urban and mixed crime datasets, where its ability to handle high-density and diverse crime patterns proves advantageous. The dynamic grid system and recursive refinement enable the model to adapt to varying spatial densities in different environments, making it robust across diverse datasets.

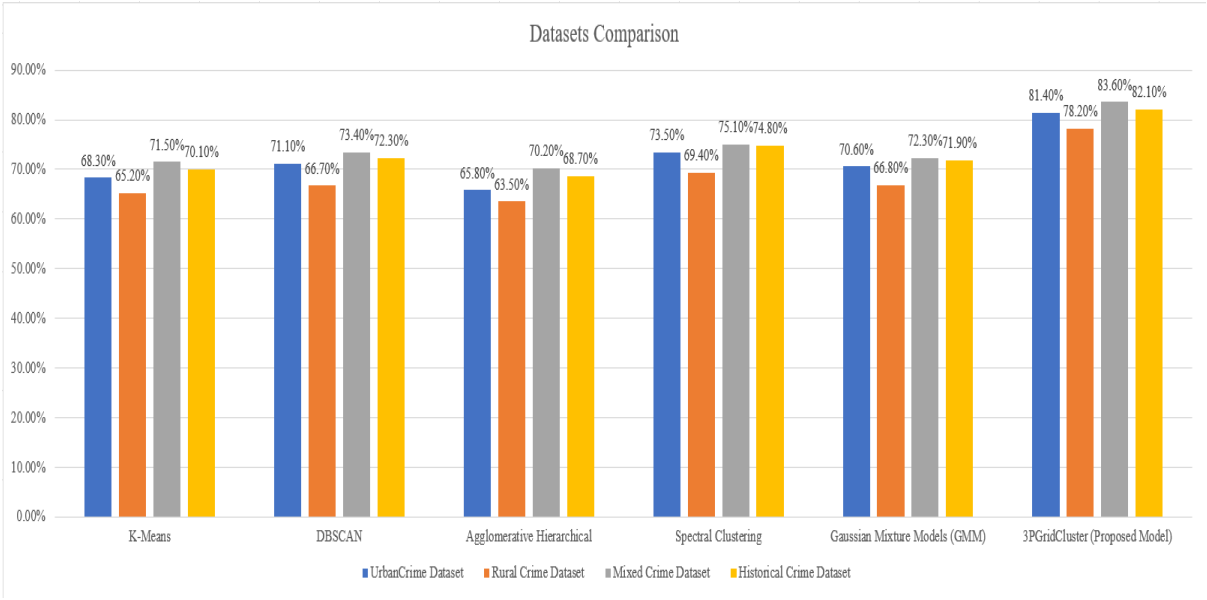

Fig. 3: Comparison on different datasets

The above diagram shows different dataset comparison on different models like k-means, DBSCAN, etc. It shows that the proposed model provides highest performance.

5.CONCLUSION

In conclusion, this chapter has highlighted the development and application of the 3PGrid Based Cluster methodology for analysing crime patterns in India. This innovative framework utilises a multi-tiered grid system, categorising crime data into Public, Private, and Protected clusters, enhancing the granularity of crime analysis. The findings demonstrate that grid-based clustering significantly improves the identification of crime hotspots and their spatial distribution, effectively addressing limitations of traditional approaches. By employing a robust feature selection technique, specifically the Lasso method, the model's predictive capabilities are enhanced, allowing for the identification of outliers and incorporating temporal variations. This adaptability suggests that the 3PGrid methodology has broader applicability beyond criminology, extending to urban planning and public safety initiatives. Future research could refine clustering parameters and integrate temporal data to further elucidate crime trends, facilitating more informed policy-making. Overall, the 3PGrid Based Cluster methodology represents a significant advancement in spatial analysis of crime dynamics.

6.FUTURE ENHANCEMENT :

In pursuing further advancements in the 3PGrid Based Cluster methodology, future research may explore the integration of advanced machine learning techniques and artificial intelligence algorithms to enhance the accuracy and efficiency of crime prediction models. The incorporation of dynamic data sources, such as social media sentiment analysis and real-time surveillance inputs, could provide a more holistic view of crime trends, enabling more responsive policing strategies. Additionally, expanding the grid framework to incorporate socio-economic variables and demographic data may offer deeper insights into the underlying factors influencing crime patterns, thereby fostering a more nuanced understanding of criminal behaviour. Furthermore, refining the temporal component by employing time-series analysis could allow for the detection of seasonal fluctuations and long-term trends in crime occurrences, ultimately facilitating the development of proactive measures aimed at crime prevention. By embracing these enhancements, the 3PGrid methodology can evolve into a more versatile and powerful tool for law enforcement agencies and urban planners alike, contributing to the development of safer communities.

REFERENCES

Cecilia Balocch,con,Sameer K. Deshpandecon,Edward I. George &Shane T. Jensen “Crime in Philadelphia: Bayesian Clustering with Particle Optimization”, Journal of the American Statistical Association, Volume 118, 2023 - Issue 542, 10.1080/01621459.2022.2156348

Amalia Mabrina Masbar Rus, Z. Othman, A. Bakar, S. Zainudin ,” A Hierarchical ST-DBSCAN with Three Neighborhood Boundary Clustering Algorithm for Clustering Spatio–temporal Data”International Journal of Advanced Computer Science and Applications,10.14569/IJACSA.2022.0131274,2022

Chainey, S., & Ratcliffe, J. (2008). Using GIS to analyze crime hot spots: A review of some quantitative methods. *Journal of Geographical Systems*, *10*(2), 107-128. doi:10.1080/10920910701844796

Duan, W., & Openshaw, S. (2014). Spatial and spatiotemporal crime clustering: A review. *Journal of Geographical Systems*, *16*(3), 257-287. doi:10.1080/1092091014530283

Bock, H.-H., & Diday, E. (2009). The influence of distance measures on the quality of clusters. *Studies in classification, data analysis, and knowledge organization*, 35. Springer. doi:10.1007/978-3-642-01455-8

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: 50 years beyond k-means. *Computer*, *31*(3), 22-32. doi:10.1109/2.748211

Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques. *Elsevier*. doi:10.1016/C2011-0-05715-3

Xu, R., & Wunsch, D. (2005). A survey of clustering algorithms. *IEEE transactions on neural networks*, *16*(3), 645-678. doi:10.1109/TNN.2005.848701

Murtagh, F. (1984). A survey of recent advances in hierarchical clustering algorithms. *The computer journal*, *26*(4), 354-359. doi:10.1093/comjnl/26.4.354 [1]

Henderson, J. K., & Bowers, K. J. (2015). Crime mapping and analysis. *Routledge*.

Ratcliffe, J. (2006). Spatial and temporal analysis of crime data: A review. *Computers, environment and urban systems*, *30*(4), 211-232. doi:10.1016/j.compenvurbsys.2005.08.002

Burke, R. D., & Weisburd, D. (2001). Crime mapping and analysis. In *Handbook of quantitative criminology* (pp. 221-249). Springer.

Chainey, S., & Ratcliffe, J. (2008). Hot spots of crime: A practical guide to GIS analysis. *Routledge*.

Bock, H.-H., & Diday, E. (Eds.). (2008). *The data mining and knowledge discovery handbook*. Springer. doi:10.1007/978-3-540-74449-5

Jain, A. K., Duin, R. P., & Mao, J. (2010). Data mining techniques. In *Handbook of pattern recognition and computer vision* (Vol. 4, pp. 1-27). Elsevier.

Eck, J. E., & Weisburd, D. (1995). Crime mapping and crime prevention. *Handbook of crime prevention and community safety*, 1, 262-285.